INTRODUCTION

The common bean (*Phaseolus vulgaris*) is a herbaceous annual plant domesticated independently in ancient Mesoamerica and the Andes, and now grown worldwide for its edible bean, popular both dry and as a green bean. The leaf is occasionally used as a leaf vegetable, and the straw is used for fodder. It is a staple food for many people due to its energy, protein, dietary fibre and minerals content (Haytowitz et al., 1981; Norton et al., 1985). Botanically, the common bean is classified as a dicotyledon. The seeds of common bean have valuable nutritional properties due to the fact that they are an important source of fibre, minerals and vitamins, as well as to their low content of fat and sodium (Sgarbieri and Whitaker, 1982). A diet including beans provides substantial health benefits, decreasing the risk of heart and renal diseases (Kabagambe et al., 2005), protecting against several cancer types (Kolonel et al., 2000) and helping in the control of overweight and obesity (Celleno et al., 2007).

The common bean containing major seed storage protein is known as the phaseolin (Bollini and Vitale, 1981; Gepts and Bliss, 1986). Phaseolin is the salt-soluble glycoprotein that accounts for some 50 per cent of the total protein in mature bean seeds. The computational packages and online servers are the current tools used in the nucleotide or protein sequence analysis and characterization. The query full length nucleotide sequence of phaseolin gene was subjected to BLAST analysis with the homologous nucleotide sequences. The sequences homologues to the phaseolin gene query sequence were retrieved on the bases of query coverage, maximum identity and e-value.

RESEARCH ARTICLE

BLAST analysis of Phaseolin gene from common bean (*Phaseolus vulgaris*)

ROSY RANI1, PANKAJ KUMAR2, JITENDER SINGH2, AMIT KUMAR SINGH1, ANCHAL RANI1, PRADEEP SHUKLA1 AND PRAGATI MISRA1*

1Sam Higginbottom Institute of Agriculture, Technology and Sciences, ALLAHABAD (U.P.) INDIA
2Sardar Vallabhbhai Patel University of Agriculture and Technology, MEERUT (U.P.) INDIA

ARTICLE INFO

Received : 17.02.2014
Revised : 04.03.2014
Accepted : 14.03.2014

Key Words :
*Phaseolus vulgaris*, Phaseolin gene, Nucleotide Database, BLAST


ABSTRACT

The common bean (*Phaseolus vulgaris*) contains major seed storage protein which is known as the phaseolin. The phaseolin is the salt-soluble glycoprotein that accounts of some 50 per cent of the total protein in mature bean seeds. The computational packages and online servers are the current tools used in the nucleotide or protein sequence analysis and characterization. The query full length nucleotide sequence of phaseolin gene was subjected to BLAST analysis with the homologous nucleotide sequences. The sequences homologues to the phaseolin gene query sequence were retrieved on the bases of query coverage, maximum identity and e-value.

*Corresponding author:
Email: pragatimisra3@rediffmail.com
analysis of biological data.

BLAST stands for Basic Local Alignment Search Tool. It was developed by the National Centre for Biotechnology Information (NCBI) in USA. Even though the name of BLAST suggests only local alignment, but in reality it can carry out both local and global alignment. The main idea behind BLAST was to feed the computer with an input sequence and produce an output when it is run against a selected database. The output is then interpreted to derive information about the input sequence. As mentioned earlier, the main purpose of using BLAST is sequence alignment. In carrying out a local alignment, BLAST breaks down an input sequence into smaller parts and compares them with the database. No gaps are introduced in local alignment in order to force the input sequence to match with the database. A match or a mismatch is indicated by the presence or absence of a vertical line between the alphabets of the input sequence and the database sequence. Unlike local alignment, sequences are guaranteed to match with the input sequence. Here, gaps are introduced whenever a mismatch is found and the input sequence is shifted by some places in order to match it with the database. One other important aspect of global alignment is that the whole sequence is taken at once and matched with the database, unlike in local alignment where an input sequence is broken down into smaller components.

MATERIAL AND METHODS

The BLAST (Basic Local Alignment Search Tool) is an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of different proteins or the nucleotides of DNA sequences. A BLAST search enables a researcher to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold. The obtained full length nucleotide sequence of *Phaseolin* gene through sequencing from YVR life sciences, Ghaziabad was subjected to BLAST analysis (BLASTX, National Center for Biotechnology Information - NCBI) with the homologous nucleotide sequences. The sequences homologues to the query sequence were retrieved on the bases of query coverage, maximum identity and e-value.

BLAST analysis for the alignment:

BLAST (Basic Local Alignment Search Tool) is an effective tool in bioinformatics. Inside the blast homepage (http://blast.ncbi.nlm.nih.gov/Blast.cgi) BLAST option was clicked. In the next page, the fasta format of this query nucleotide sequence was uploaded and selects the nucleotide database and finally clicked on BLAST option. After a waiting time the BLAST result appeared.

RESULTS AND DISCUSSION

The full length nucleotide sequence of *Phaseolin* gene was subjected to BLAST analysis (Table 1) with earlier reported phaseolin genes, which were retrieved from the NCBI database. The ten sequences were showed high range of query coverage with low e-value and maximum identity were used for BLAST analysis. These sequences showed maximum identity i.e. from 79 per cent to 100 per cent and wide range of query coverage i.e. from 27 per cent to 98 per cent with low e-value. The sequences *Phaseolus vulgaris* mRNA for beta-type phaseolin precursor (accession no. X03004.1), *Phaseolus vulgaris* mRNA for alpha-type phaseolin precursor (accession no. X02980.1), *Phaseolus vulgaris* Sanilac clone 2-13 phaseolin (Phs) mRNA, complete cds (accession no. U01132.1) and *Phaseolus vulgaris* Sanilac clone 1-12 phaseolin (Phs)...

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>max score</th>
<th>total score</th>
<th>query coverage</th>
<th>e value</th>
<th>max identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>X52626.1</td>
<td><em>Phaseolus vulgaris</em> gene for alpha-phaseolin</td>
<td>3253</td>
<td>3253</td>
<td>98%</td>
<td>0.0</td>
<td>97%</td>
</tr>
<tr>
<td>J01263.1</td>
<td><em>Phaseolus vulgaris</em> beta-type phaseolin storage protein gene, complete cds</td>
<td>3127</td>
<td>3451</td>
<td>98%</td>
<td>0.0</td>
<td>99%</td>
</tr>
<tr>
<td>V01163.1</td>
<td>Part of the gene for phaseolin in <em>Phaseolus vulgaris</em> (bean)</td>
<td>1450</td>
<td>1450</td>
<td>46%</td>
<td>0.0</td>
<td>96%</td>
</tr>
<tr>
<td>X03004.1</td>
<td><em>Phaseolus vulgaris</em> mRNA for beta-type phaseolin precursor</td>
<td>584</td>
<td>2558</td>
<td>72%</td>
<td>9e-163</td>
<td>100%</td>
</tr>
<tr>
<td>X02980.1</td>
<td><em>Phaseolus vulgaris</em> mRNA for alpha-type phaseolin precursor</td>
<td>564</td>
<td>2296</td>
<td>70%</td>
<td>1e-156</td>
<td>100%</td>
</tr>
<tr>
<td>U01132.1</td>
<td><em>Phaseolus vulgaris</em> Sanilac clone 2-13 phaseolin (Phs) mRNA, complete cds</td>
<td>564</td>
<td>2363</td>
<td>73%</td>
<td>1e-156</td>
<td>100%</td>
</tr>
<tr>
<td>U01131.1</td>
<td><em>Phaseolus vulgaris</em> Sanilac clone 1-12 phaseolin (Phs) mRNA, complete cds</td>
<td>562</td>
<td>2525</td>
<td>73%</td>
<td>4e-151</td>
<td>100%</td>
</tr>
<tr>
<td>U01121.1</td>
<td><em>Phaseolus lunatus</em> phaseolin (Phs) mRNA, complete cds</td>
<td>492</td>
<td>1201</td>
<td>42%</td>
<td>6e-135</td>
<td>95%</td>
</tr>
<tr>
<td>AM905848.1</td>
<td>Vigna unguiculata partial vicilin gene, exons 1-6</td>
<td>545</td>
<td>545</td>
<td>42%</td>
<td>4e-151</td>
<td>79%</td>
</tr>
<tr>
<td>DQ538335.1</td>
<td>Vigna radiata 8S globulin beta isoform precursor mRNA, complete CDS</td>
<td>462</td>
<td>839</td>
<td>27%</td>
<td>4e-126</td>
<td>95%</td>
</tr>
</tbody>
</table>
mRNA, complete cds (accession no. U01131.1) showed 100 per cent identity with the query sequence of Phaseolin gene. These sequences showed 70 per cent to 73 per cent query coverage and low e-value. These data indicated that these sequences rely between the query sequence of Phaseolin gene. The Phaseolus vulgaris beta-type phaseolin storage protein gene, complete cds (accession no J01263.1) showed 99 per cent identity with the query coverage of 98 per cent with zero e-value. The Phaseolus vulgaris gene for alpha-phaseolin (accession no. X52626.1) has 97 per cent identity with query sequence and 98 per cent query coverage with zero e-value. The part of the gene for Phaseolin in Phaseolus vulgaris (bean) (accession no. V01163.1) showed 96 per cent identity with 46 per cent query coverage and zero e-value whereas Phaseolus lunatus phaseolin (Phs) mRNA, complete cds (accession no. U01121.1) and Vigna radiata 8S globulin beta isoform precursor mRNA, complete CDS (accession no. DQ538335.1) showed 95 per cent identity with 42 per cent and 27 per cent query coverage, respectively and low e-value. The Vigna unguiculata partial vicilin gene, exons 1-6 (accession no. AM905848.1) showed 79 per cent identity with low e-value and query coverage of 42 per cent on BLAST analysis with query sequence of Phaseolin gene (Table 1).

BLAST (Basic Local Alignment Search Tool) (Altschul et al., 1990, Altschul et al., 1997) is a widely used bioinformatics application for rapidly querying nucleotide (DNA) and protein sequence databases. Given a query sequence, the goal is to find the most similar sequences from a large database. This has applications in the identification of functions and structures of unknown sequences or understanding the evolutionary origin of DNA or protein query sequences. Due to its high importance, BLAST also finds a place in almost all the bioinformatics solutions.

REFERENCES


